

PRACA DOMOWA 2

**WIKTOR
BUDZIŃSKI**

PRACA DOMOWA

Przeprowadź analizę wybranego zbioru danych wykorzystując metody poznane na zajęciach

- Wybrany zbiór powinien być mikro
 - Tzn. obserwowana jednostka to konsument / gospodarstwo domowe / firma itd.
 - Ewentualnie mała jednostka przestrzenna, np. gmina
- Przeprowadzona analiza powinna **mieć jakiś cel**
 - Badanie interesującego związku między zmienną objaśnianą a objaśniającą
- Zmienna objaśniana musi być jednego z 4 typów:
 - Binarna
 - Uporządkowana
 - Wielomianowa
 - Liczebności zdarzeń
- Zbiór danych powinien mieć co najmniej 300 obserwacji oraz co najmniej 5 zmiennych objaśniających
 - Jeśli zbiór danych jest bardzo duży, co sprawia, że modele się liczą bardzo długo, można pracować na jego podpróbie
 - Powinna być ona losowa, przed losowaniem należy ustawić ziarno aby wyniki były możliwe do powtórzenia
 - W STATA można wykorzystać komendy:
 - `sample 60` (losuje 60% obserwacji)
 - `sample2 60, cluster(id)` (losuje 60% obserwacji, ale bierze pod uwagę, że kilka wierszy może należeć do tego samego respondenta – niezbędne przy danych wielomianowych)

PRACA DOMOWA: POLECENIE

1. Przeprowadź wstępną analizę jakościową danych
 - A. Np. analiza graficzna, korelacje itd.
2. Przeprowadź analizę zmiennej objaśnianej przy pomocy odpowiedniego modelu
 - A. Przeprowadź diagnostykę modelu
 - B. Popraw formę funkcyjną modelu
 - a) Przede wszystkim skupiając się na zmiennych istotnych dla celu analizy
 - b) Do oceny formy funkcyjnej wykorzystaj formalny test statystyczny oraz miary dopasowania
 - C. Wykorzystaj rozszerzenia danego modelu omawiane na zajęciach
 - D. Wykorzystaj testy Walda i LR do przetestowania pewnej hipotezy prostej i złożonej
 - E. Zinterpretuj wyniki najlepszego wg Ciebie modelu
 - a) Zarówno jakościowo jak i ilościowo
3. Wyciągnij wnioski z przeprowadzonej analizy
 - A. Wnioski powinny dotyczyć celu analizy i wynikać z policzonych modeli ekonometrycznych

ZASADY ODDANIA

- Praca do wykonania w parach
- **Do 12.05** proszę o informację jaki zbiór danych chcą Państwo analizować
 - Z dodatkową informacją o zmiennej objaśnianej, zmiennych objaśniających oraz celu analizy
 - Proszę podać źródło danych
 - **Kto pierwszy ten lepszy**, każda grupa pracuje na innym zbiorze danych. Zachęcam do wcześniejszych zgłoszeń.
- **Do 26.05** proszę wysłać pracę domową na maila wiktor_budzinski@uw.edu.pl
 - Zbiór danych w formacie .dta
 - Kody wykorzystane do analizy (wraz z komentarzami) w formacie .do
 - Raport z przeprowadzonej analizy w formacie .pdf
 - Krótki opis celu analizy
 - Krótki opis zbioru danych z analizą jakościową
 - Wyniki przeprowadzonej analizy ilościowej
 - Nie trzeba tutaj opisywać wszystkich policzonych modeli. Natomiast wszystkie modele powinny być w do-file'u tak żebym mógł przeanalizować Państwa tok rozumowania
 - Krótkie podsumowanie

PRZYKŁADOWE ZBIORY DANYCH

Przykładowe zbiory dla zmiennej binarnej:

- Samozatrudnienie wśród kobiet
 - [Journal of Applied Econometrics Data Archive: Data Directory \(queensu.ca\)](#)
- Ile skłonne są zapłacić gospodarstwa domowe za poprawę w dostawach prądu
 - [Households' valuation of power outages in major cities of Ethiopia: An application of stated preference methods – ScienceDirect](#)
 - Dane na dole strony w „Appendix C. Supplementary data.” Możliwe, że potrzebny dostęp przez stronę uniwersytetu.
- ~~Co wpływa na to, że pacjent zostanie przyjęty do szpitala?~~
 - [Data and Code for: Time Dependency in Physician Decision-Making \(openicpsr.org\)](#)
 - Raczej duży zbiór danych

Przekreślony zbiór danych oznacza, że ktoś już zgłosił, że chce na nim pracować

PRZYKŁADOWE ZBIORY DANYCH

Przykładowe zbiory dla zmiennej uporządkowanej:

- Co wpływa na poziom stresu?
 - [Time Use and Stress \(openicpsr.org\)](https://openicpsr.org/openicpsr/data/stress/time-use)
- Co wpływa na szczęście?
 - [Replication data for: The Paradox of Declining Female Happiness \(openicpsr.org\)](https://openicpsr.org/openicpsr/data/happiness/replication)

Przekreślony zbiór danych oznacza, że ktoś już zgłosił, że chce na nim pracować

PRZYKŁADOWE ZBIORY DANYCH

Przykładowe zbiory dla zmiennej wielomianowej:

- ~~Wpływ eko-etykiet na zakup piwa~~
 - [Consumer WTP for Sustainability Attributes in Beer - Mendeley Data](#)
- Preferencje odnośnie zakazu wjazdu samochodów do centrum miasta
 - [Data for: Investigating People's Preferences for Car-Free City Centers: A Discrete Choice Experiment - Mendeley Data](#)
- Preferencje odnośnie urodzeń
 - [Journal of Applied Econometrics Data Archive: Data Directory \(queensu.ca\)](#)

Przekreślony zbiór danych oznacza, że ktoś już zgłosił, że chce na nim pracować

PRZYKŁADOWE ZBIORY DANYCH

Przykładowe zbiory dla zmiennej wielomianowej:

- Preferencje odnośnie podziemnych linii wysokiego napięcia
 - [Heterogeneous public preferences for undergrounding high-voltage power transmission lines: The case of Seoul metropolitan area in South Korea – ScienceDirect](#)
 - Dane na dole strony w „Appendix F. Supplementary data.” Możliwe, że potrzebny dostęp przez stronę uniwersytetu.
- Preferencje odnośnie soków GMO
 - [Investigating Consumer Stated Preferences for Gene-edited Orange Juice – The Influence of Behavioral Traits - Harvard Dataverse](#)
 - Duży zbiór danych
- Ewentualnie więcej zbiorów danych:
 - [alvarogutierrez/TheDiscreteChoiceDataBank: Collection of publicly available Discrete Choice Datasets \(github.com\)](#)

Przekreślony zbiór danych oznacza, że ktoś już zgłosił, że chce na nim pracować

PRZYKŁADOWE ZBIORY DANYCH

Przykładowe zbiory dla zmiennej dotyczącej liczebności zdarzeń:

- Chodzenie na siłownię
 - [Replication data for: Paying Not to Go to the Gym \(openicpsr.org\)](https://openicpsr.org/)
- Zakupy fotowoltaiki
 - [Heterogeneity in the adoption of photovoltaic systems in Flanders – ScienceDirect](#)
 - Dane na dole strony w „Appendix A. Supplementary data.” Możliwe, że potrzebny dostęp przez stronę uniwersytetu.

Przekreślony zbiór danych oznacza, że ktoś już zgłosił, że chce na nim pracować